# UN Committee of Experts on Big Data and Data Science for Official Statistics

## Revised Mandate

Ronald Jansen, Assistant Director

United Nations Statistics Division

Email: jansen1@un.org ;
BigData@un.org

United Nations Statistics Division

# UN Committee of Experts on Big Data and Data Science for Official Statistics

Created in March 2014 by the UN Statistical Commission (annually reporting)

Inter-governmental body with 31 countries and 16 international organizations

Collaboration of more than 400 experts from all stakeholder communities

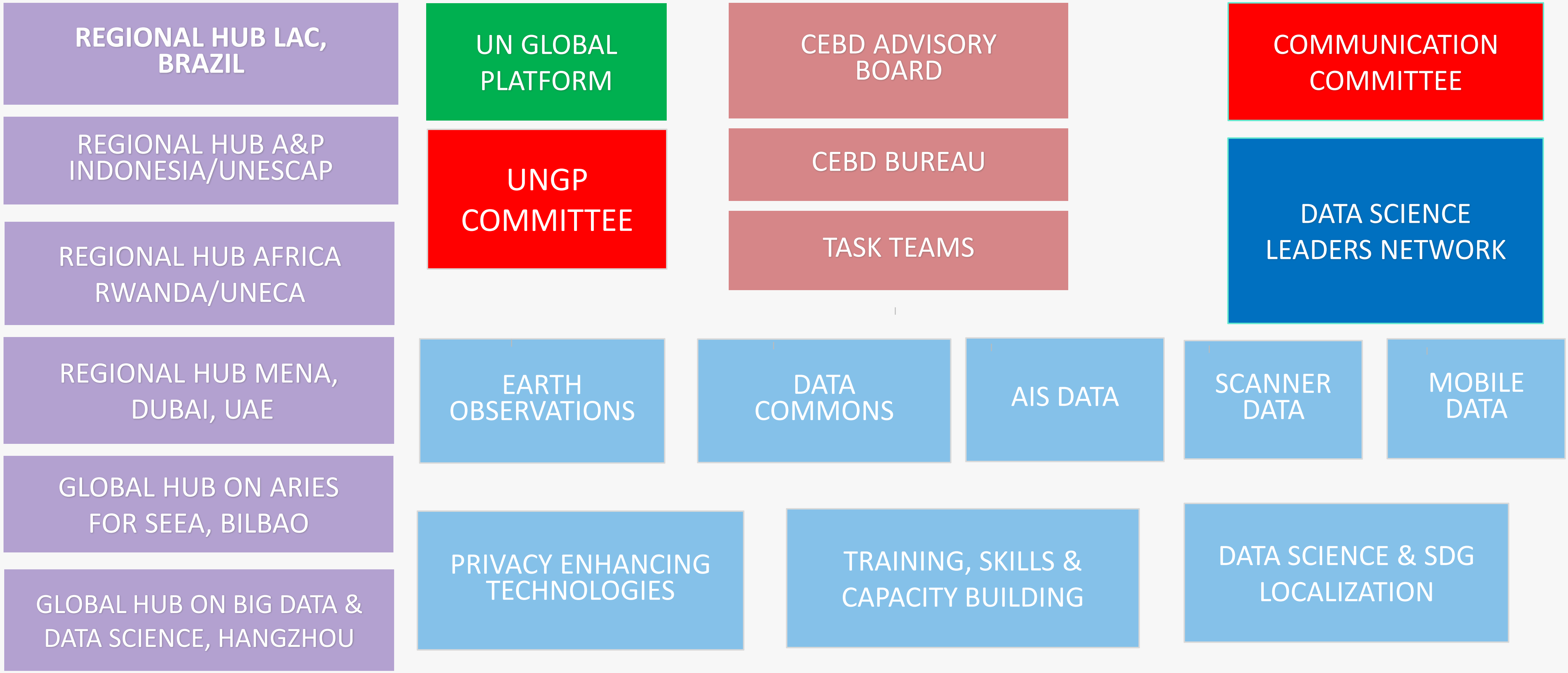
United Nations
BigData

# Mandate (Decision 45/110 – 2014)

- Provide strategic vision of a global programme on Big Data for official statistics;
- Promote practical use of sources of Big Data and find solutions for
    - Methodological issues,
    - Legal issues of access to data sources;
    - Privacy issues
    - Data security issues;
    - Cost benefit analysis
- Promote capacity building
- Foster Communication and Advocacy
- Build Public Trust

United Nations BigData

# UN Committee of Experts on Big Data and Data Science for official statistics



REGIONAL HUB LAC, BRAZIL

REGIONAL HUB A&P INDONESIA/UNESCAP

REGIONAL HUB AFRICA RWANDA/UNECA

REGIONAL HUB MENA, DUBAI, UAE

GLOBAL HUB ON ARIES FOR SEEA, BILBAO

GLOBAL HUB ON BIG DATA & DATA SCIENCE, HANGZHOU

UN GLOBAL PLATFORM

UNGP COMMITTEE

CEBD ADVISORY BOARD

CEBD BUREAU

TASK TEAMS

COMMUNICATION COMMITTEE

DATA SCIENCE LEADERS NETWORK

EARTH OBSERVATIONS

DATA COMMONS

AIS DATA

SCANNER DATA

MOBILE DATA

PRIVACY ENHANCING TECHNOLOGIES

TRAINING, SKILLS & CAPACITY BUILDING

DATA SCIENCE & SDG LOCALIZATION

# Organization of UNCEBD

## Task Teams:

- EO for agriculture statistics
- Mobile phone data for official stats
- Scanner data for price statistics
- AIS data for transport statistics
- Training, capacity development
- Data commons / Data spaces
- Privacy Enhancing Technologies
- Data Science and SDG localization

## Committees:

- Data Science Leaders Network
- UNGP Committee
- Communication Committee

## Hubs:

- Global Hubs (China & Spain)
- Regional Hubs (Brazil, Indonesia, Rwanda & UAE)

# Manuals

MPD for official statistics:

- Tourism
- Migration
- Disaster and Displacement
- Transport
- Information Society

EO for official statistics:

- 2017 manual
- 2022 manual

UN PET Guide (2023)

Scanner data for price stats (2024)

# E-Learning courses

# Projects

## On-going Projects on UNGP:

- Privacy-preserving data science
- Vessel tracking data (AIS)
- Climate & health indicators
- Statistical data portals (.Stat)
- Modernization of UN Data
- New trade data processing tools
- E-learning courses
- Data4Now

## Concluded Projects:

- Using satellite imagery & machine learning to create modern crop maps in Senegal
- Concept & SDG extraction using semantic web technologies

UN Big Data

United Nations | Department of Economic and Social Affairs

# Projects

## On-going Projects:

- Privacy-preserving data science
- Vessel tracking data (AIS)
- Climate & health indicators
- Statistical data portals (.Stat)
- Modernization of UN Data
- New trade data processing tools
- E-learning courses
- Data4Now

## Concluded Projects:

- Using satellite imagery & machine learning to create modern crop maps in Senegal
- Concept & SDG extraction using semantic web technologies

UNBigData

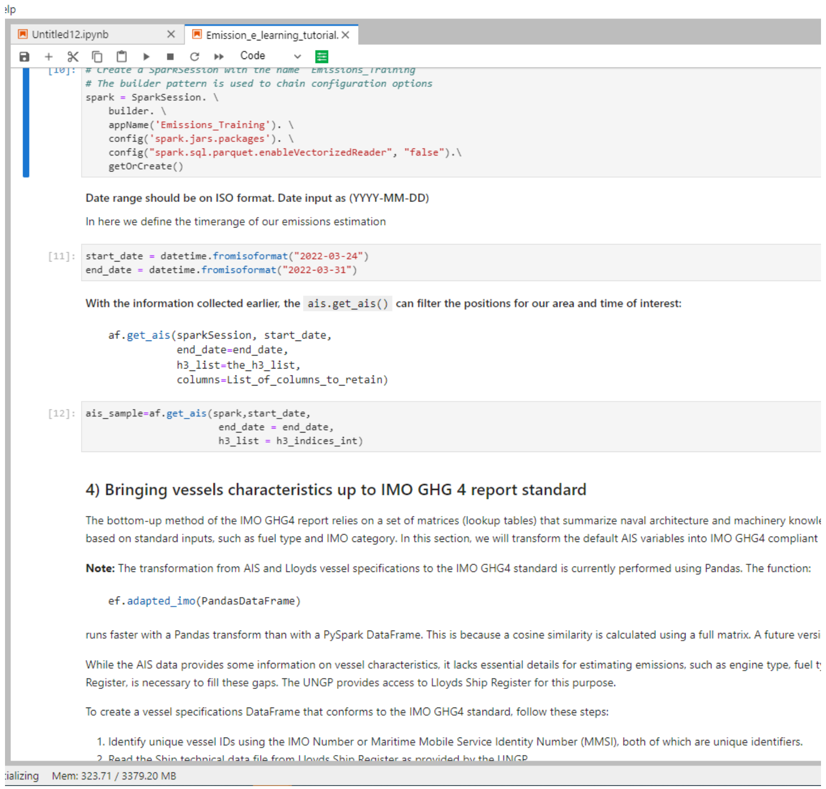
United Nations | Department of Economic and Social Affairs

# Vessel-tracking data (AIS) on UNGP

- ❑ **Audience:** Global Statistical Community
- ❑ **Goal:** Facilitate access to TB of AIS vessel location data in a cost-effective, cloud-based computing environment
- ❑ We built **data pipelines** to continuously ingest AIS data **for data scientists** to run their own solutions, resulting in 15+ research papers plus several data platforms (such as IMF PortWatch), with data going back to 2018.

# The AIS Service on the UN Global Platform

- **Platform-as-a-service:** users should be able to build their own solutions (like *PortWatch*)
- Built using modern **cloud-native** technologies (K8s, spot instances, serverless)
- Extensive **partnerships** with NSOs in user experience & peer review of technology architectures
- In-house **operations** and **engineering**
- Users **prototype solutions** in Notebook environments
- Also provides a **remote data processing interface** to execute pipelines remotely for remote execution by partners


elp
Untitled12.ipynb
Emission_e_learning_tutorial.
Code
ializing    Mem: 323.71 / 3379.20 MB

```
[10]: # Create a SparkSession with the name  Emissions_Training
      # The builder pattern is used to chain configuration options
      spark = SparkSession. \
          builder. \
          appName('Emissions_Training'). \
          config('spark.jars.packages'). \
          config("spark.sql.parquet.enableVectorizedReader", "false").\
          getOrCreate()
```

**Date range should be on ISO format. Date input as (YYYY-MM-DD)**

In here we define the timerange of our emissions estimation

```
[11]: start_date = datetime.fromisoformat("2022-03-24")
      end_date = datetime.fromisoformat("2022-03-31")
```

**With the information collected earlier, the `ais.get_ais()` can filter the positions for our area and time of interest:**

```
af.get_ais(sparkSession, start_date,
           end_date=end_date,
           h3_list=the_h3_list,
           columns=List_of_columns_to_retain)
```

```
[12]: ais_sample=af.get_ais(spark,start_date,
                            end_date = end_date,
                            h3_list = h3_indices_int)
```

### 4) Bringing vessels characteristics up to IMO GHG 4 report standard

The bottom-up method of the IMO GHG4 report relies on a set of matrices (lookup tables) that summarize naval architecture and machinery knowle based on standard inputs, such as fuel type and IMO category. In this section, we will transform the default AIS variables into IMO GHG4 compliant

**Note:** The transformation from AIS and Lloyds vessel specifications to the IMO GHG4 standard is currently performed using Pandas. The function:

```
ef.adapted_imo(PandasDataFrame)
```

runs faster with a Pandas transform than with a PySpark DataFrame. This is because a cosine similarity is calculated using a full matrix. A future versi

While the AIS data provides some information on vessel characteristics, it lacks essential details for estimating emissions, such as engine type, fuel t Register, is necessary to fill these gaps. The UNGP provides access to Lloyds Ship Register for this purpose.

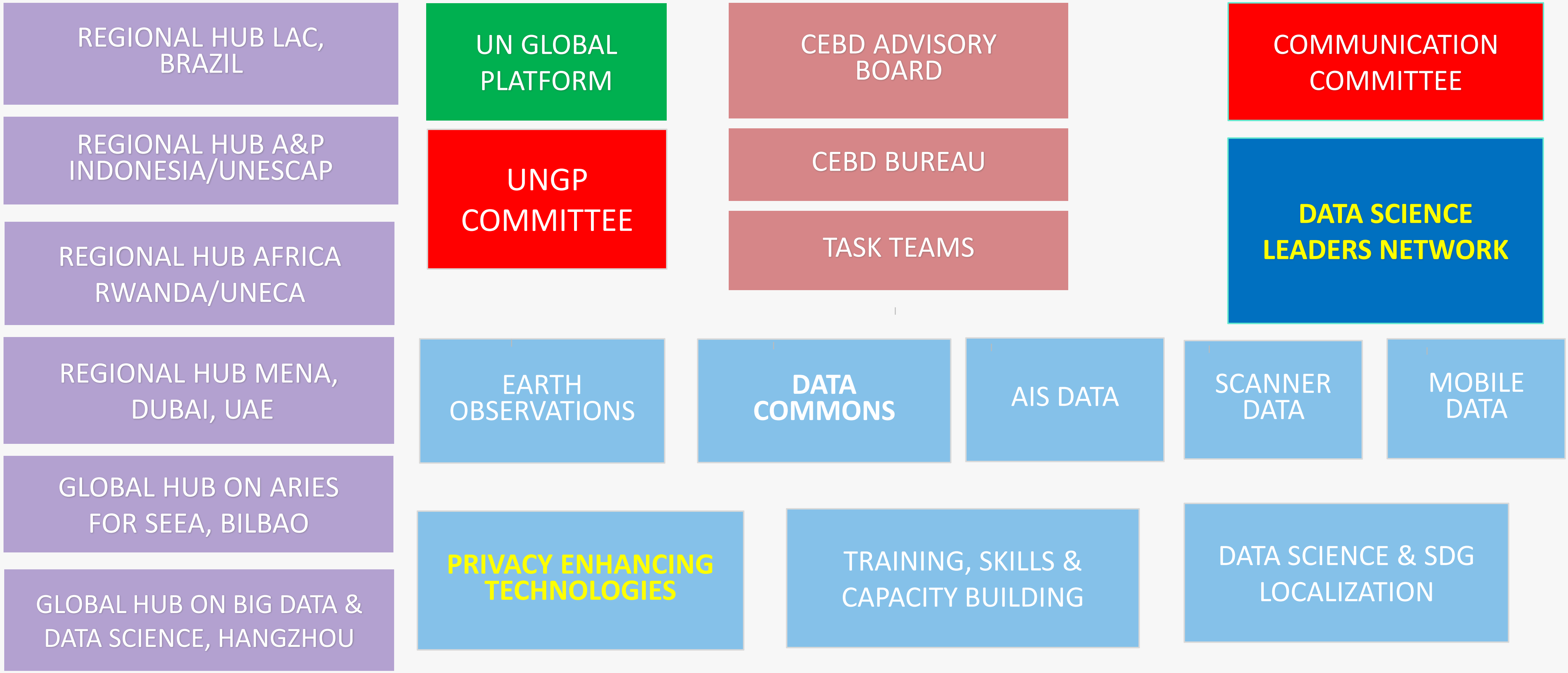
To create a vessel specifications DataFrame that conforms to the IMO GHG4 standard, follow these steps:

1. Identify unique vessel IDs using the IMO Number or Maritime Mobile Service Identity Number (MMSI), both of which are unique identifiers.
2. Read the Ship technical data file from Lloyds Ship Register as provided by the UNGP

# UNCEBD Revised Mandate 2025 (draft)

- To provide a strategic vision, direction and coordination for a global programme on emerging new statistical methodologies and technologies, such as **AI, data science and the use of big data and other alternative data sources** for the improvement of official statistics
- To promote practical use of AI, data science and the use of big data and other alternative data sources, while building on existing precedents and finding solutions for the many existing challenges.
- To promote strategic relationships with private sector, geospatial community, academia and other public sector institutes to ensure better access to data and responsible use of AI
- To promote data governance, AI governance, data stewardship and open data policies for better access and use of data.

# UN Committee of Experts on Big Data and Data Science for official statistics



REGIONAL HUB LAC, BRAZIL

REGIONAL HUB A&P INDONESIA/UNESCAP

REGIONAL HUB AFRICA RWANDA/UNECA

REGIONAL HUB MENA, DUBAI, UAE

GLOBAL HUB ON ARIES FOR SEEA, BILBAO

GLOBAL HUB ON BIG DATA & DATA SCIENCE, HANGZHOU

UN GLOBAL PLATFORM

UNGP COMMITTEE

CEBD ADVISORY BOARD

CEBD BUREAU

TASK TEAMS

COMMUNICATION COMMITTEE

**DATA SCIENCE LEADERS NETWORK**

EARTH OBSERVATIONS

**DATA COMMONS**

AIS DATA

SCANNER DATA

MOBILE DATA

**PRIVACY ENHANCING TECHNOLOGIES**

TRAINING, SKILLS & CAPACITY BUILDING

DATA SCIENCE & SDG LOCALIZATION

# Data Science Leaders Network

- ❑ Playbook
- ❑ Data Science for Statistical and Transport Communities
  - ▪ https://unstats.un.org/bigdata/events/2024/transport-seminar/
- ❑ AI and Data Science for Economic Statistics
  - ▪ 1ST Webinar – 7 Nov 2024
  - ▪ 2nd Webinar – 12 Dec 2024
  - ▪ International webinar – 20-22 Jan 2025

UNBigData

United Nations | Department of Economic and Social Affairs

# Privacy Enhancing Technologies – PETs

❑ PET Lab – Can PETs become the standard for data access?

❑ UN PET Guide

❑ Special Issue (March 2025) of SJ of IAOS

- Remote Data Science
- Risk Assessment and PET help
- Private Set Intersection (PSI) – NSO data sharing
- Introduction – Data Governance, GDC and PETs